The Pace of Prosodic Phrasing Couples the Listener's Cortex to the Reader's Voice

Mathieu Bourguignon,[1] Xavier De Tiège,[1]* Marc Op de Beeck,[1] Noémie Ligot,[1] Philippe Paquier,[1] Patrick Van Bogaert,[1] Serge Goldman,[1] Riitta Hari,[2] and Veikko Jousmäki[1,2]

[1]*Laboratoire de Cartographie Fonctionnelle du Cerveau, ULB-Hôpital Erasme, Université Libre de Bruxelles, 1070 Brussels, Belgium*
[2]*Brain Research Unit, Low Temperature Laboratory, Aalto University School of Science, Espoo, Finland*

◆━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━◆

Abstract: We studied online coupling between a reader's voice and a listener's cortical activity using a novel, ecologically valid continuous listening paradigm. Whole-scalp magnetoencephalographic (MEG) signals were recorded from 10 right-handed, native French-speaking listeners in four conditions: a female (Exp1f) and a male (Exp1m) reading the same text in French; a male reading a text in Finnish (Exp 2), a language incomprehensible for the subjects, and a male humming Exp1 text (Exp 3). The fundamental frequency ($f0$) of the reader's voice was recorded with an accelerometer attached to the throat, and coherence was computed between $f0$ time-course and listener's MEG. Similar levels of right-hemisphere-predominant coherence were found at ~0.5 Hz in Exps 1–3. Dynamic imaging of coherent sources revealed that the most coherent brain regions were located in the right posterior superior temporal sulcus (pSTS) and posterior superior temporal gyrus (pSTG) in Exps 1–2 and in the right supratemporal auditory cortex in Exp 3. Comparison between speech rhythm and phrasing suggested a connection of the observed coherence to pauses at the sentence level both in the spoken and hummed text. These results demonstrate significant coupling at ~0.5 Hz between reader's voice and listener's cortical signals during listening to natural continuous voice. The observed coupling suggests that voice envelope fluctuations, due to prosodic rhythmicity at the phrasal and sentence levels, are reflected in the listener's cortex as rhythmicity of about 2-s cycles. The predominance of the coherence in the right pSTS and pSTG suggests hemispherical asymmetry in processing of speech sounds at sub-sentence time scales. *Hum Brain Mapp 34:314–326, 2013.* © 2012 Wiley Periodicals, Inc.

Key words: speech processing; magnetoencephalography; coherence; superior temporal sulcus; superior temporal gyrus; prosody; sentence; temporal integration window

◆━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━◆

INTRODUCTION

Speech communication represents a major component of human social interaction. Successful conversation requires mutual understanding that relies on fast neuronal integration of time-varying linguistic and para-linguistic information. Indeed, speech conversation not only requires the integration of connected verbal information but also depends on decoding a large amount of nonverbal information, such as the other person's identity, affective, and cognitive states [Belin et al., 2004; Frazier et al., 2006; Hari and Kujala, 2009; Kriegstein and Giraud, 2004; Lattner et al., 2005; Rosen, 1992; Scott, 2008; von Kriegstein et al., 2003, 2010]. Much of the nonverbal information is carried by the speaker's voice, gestures, and other nonverbal cues, such as pauses and silences [Belin et al., 2004; Hari and Kujala, 2009; Kriegstein and Giraud, 2004; Lattner et al., 2005; Scott, 2008; von Kriegstein et al., 2010]. Participants engaged in a conversation use both verbal and nonverbal cues to manage the smooth flow of talking in terms of who speaks, when, for how long, and about what [Sabbagh, 1999; Wilson and Wilson, 2005].

Previous functional brain imaging results have shown that the linguistic and para-linguistic information conveyed by the voice during speech communication are processed in specialized neuronal networks [Belin et al., 2004]. Although both primary and nonprimary auditory areas are involved in speech perception, right-hemisphere temporal-lobe structures, and especially regions along the right superior temporal sulcus (STS), are involved in the processing of para-linguistic information in voices, such as affective cues and identity information [Belin et al., 2004; Hickok and Poeppel, 2007; Lattner et al., 2005; Scott, 2008; von Kriegstein et al., 2010].

Up to now, most studies addressing the neural basis of human speech communication have used highly controlled vocal or verbal stimuli. Such approaches are surely appropriate for studying the neuronal underpinnings of specific aspects of speech comprehension and production. However, evolving toward more naturalistic speech communication paradigms would help to further characterize the neurophysiological correlates of speech in everyday interactional situations [Hari and Kujala, 2009]. Moving forward with this concept, Stephens et al. [2010] used functional magnetic resonance imaging (fMRI) to demonstrate spatiotemporal coupling between a speaker's and a listener's cortical activity during listening to natural speech; the coupling diminished when the listener did not understand the spoken language. Significant coupling occurred in language-related as well as extralinguistic areas, such as precuneus, orbitofrontal, and prefrontal cortices. The listener's brain activity usually followed the speaker's brain activity with a delay of a few seconds, with varying delays to different coupled brain regions, but it sometimes even preceded the speaker's brain activity. However, the temporal resolution of fMRI is suboptimal to

fully characterize the cortical temporal hierarchy of speech processing in fleeting interactional situations. Indeed, numerous psychophysiological and neurophysiological findings suggest that most speech and nonspeech signals are processed on different timescales of less than a second within segregated nonprimary auditory brain areas [Abrams et al., 2008; Poeppel et al., 2008]. In particular, left-hemisphere nonprimary auditory areas appear to preferentially process information on short (20–50 ms) timescales, while the homologous regions in the right hemisphere process information within longer (150–250 ms) time windows [Abrams et al., 2008; Poeppel et al., 2008].

Time-sensitive neurophysiological techniques, such as magnetoencephalography (MEG) and scalp or intracranial electroencephalography (EEG), represent attractive methods to investigate brain processes that unfold at millisecond level. Previous studies using these techniques have revealed coupling between speech temporal envelope and auditory-cortex activity during word and sentence listening [Abrams et al., 2008; Ahissar et al., 2001; Aiken and Picton, 2008; Nourski et al., 2009; Suppes et al., 1997, 1998] and have suggested that the right-hemisphere auditory cortex encodes temporal information of speech at 4–8 Hz that is mainly the speech envelope that represents syllable patterns [Abrams et al., 2008]. These studies typically used recorded words or single sentences produced in conversational, compressed, or distorted speech modes [Abrams et al., 2008; Ahissar et al., 2001; Aiken and Picton, 2008; Nourski et al., 2009; Suppes et al., 1997, 1998]. Investigating the coupling between speech signals and brain activity in everyday verbal communication situations would potentially bring further understanding of the neural basis of human speech processing. Therefore, to further characterize the cortical basis of auditory processing in ecologically valid continuous listening conditions, we developed a novel coherence analysis method to assess the degree of coupling between a reader's voice and a listener's cortical MEG signals. This "corticovocal coherence" analysis was specifically designed (1) to search for coupling between the time course of the reader's voice fundamental frequency ($f0$, i.e., the voice signal band-pass filtered around $f0$) and the listener's cortical MEG signals, (2) to identify, without any *a priori* hypothesis, the frequencies at which this coupling occurs, (3) to characterize the underlying coherent neuronal network and possible hemispheric lateralization, and (4) to study differences between speech and non-speech sounds. This type of analysis has been previously used in human neurosciences to search for significant coupling between cortical and muscular signals (corticomuscular coherence) [Baker, 2007; Conway et al., 1995; Salenius et al., 1997], cortical activity, and the kinematics of voluntary movements (corticokinematic coherence) [Bourguignon et al., 2011; Jerbi et al., 2007] and cortical network activities (corticocortical coherence) [Gross et al., 2001].
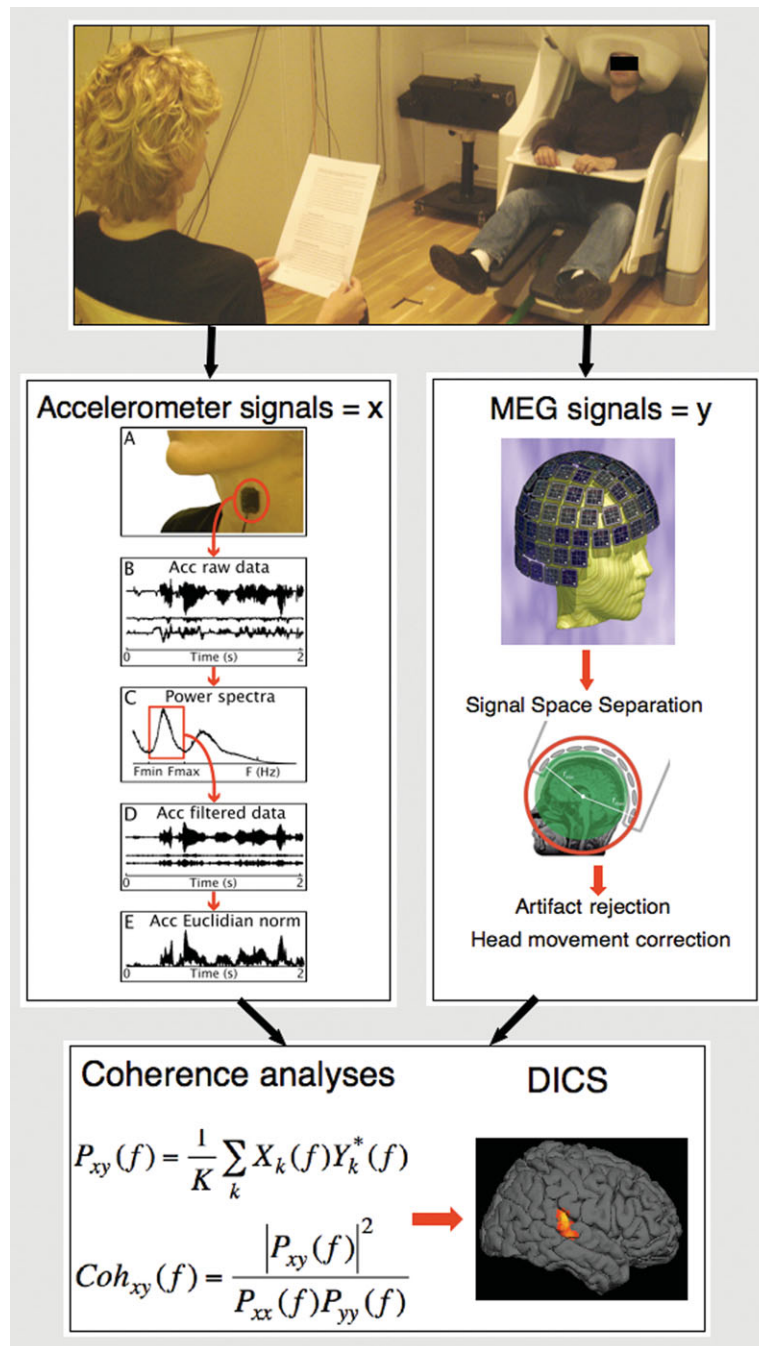
Figure 1.

Principle of the applied coherence method. *Top:* Magnetoencephalographic (MEG) signals of a subject listening to a female reading. *Middle left:* The reader's voice fundamental frequency (f0; **C**) was recorded using a three-axis accelerometer (**A** and **B**) attached to the throat. The accelerometer signals were band-pass-filtered around f0 (**C** and **D**) and then merged using the Euclidian norm (**E**). *Middle right:* MEG signals were preprocessed using signal space separation method to subtract magnetic interferences and to correct for head movements. *Bottom:* Cortico-vocal coherence was computed to determine the degree of coupling between reader's f0 time course and listener's MEG signals. Brain regions showing significant coherence were identified using dynamic imaging of coherent sources (DICS).

SUBJECTS AND METHODS

Subjects

Ten right-handed and native French-speaking healthy subjects (range, 21–38 years; mean age: 25 years; 5 females, and 5 males) without any history of neuropsychiatric disease were included in this study. Handedness was assessed by Edinburgh Handedness Inventory. The study had prior approval by the ULB-Hôpital Erasme Ethics Committee. The subjects gave informed consent before participation.

Experimental Paradigm

Cortical neuromagnetic signals were recorded using a whole-scalp MEG, while the subjects listened to a text read by one of the investigators sitting 2 m in front of the subject inside the magnetically shielded room. Alive natural voices rather than recorded voices were used as stimuli, because the main aim was to investigate auditory processing in natural listening situations. Reader's vocal activity was recorded, time-locked to MEG signal acquisition, with a three-axis accelerometer (ADXL330 iMEMS Accelerometer, Analog Devices, Norwood, MA) attached to the left side of the reader's throat (see Fig. 1). Accelerometers can accurately measure vocal activity and, in particular, the voice $f0$ [Hillman et al., 2006; Lindström et al., 2009, 2010; Orlikoff, 1995]. Compared with microphones, accelerometers are not sensitive to environmental sounds and are therefore well suitable for voice assessment [Lindström et al., 2009]. This aspect was particularly important in the current study, because we wanted to avoid any influence of environmental sounds on the coherence analysis; potential external sound sources were the background noise of the MEG electronics, the nonspeech sounds produced by the reader (page turning, etc.), and the noise produced by the subjects themselves. We concentrate on reader's $f0$ as a highly relevant voice signal for speech processing. Moreover, although $f0$ and spectral formants are correlated to some extent, they may be processed in functionally distinct brain areas [Lattner et al., 2005]. Therefore, in an attempt to be more specific in our coherence analyses, we focused on $f0$ signals as recorded by the accelerometer.

During the measurements, subjects were asked to fixate the gaze at a point in the magnetically shielded room to avoid any gaze contact with the reader. In the first experiment (Exp 1), a native French-speaking female (NL, Exp1f) and native French-speaking male (XDT, Exp1m) read for 5 min a French text (informed consent of an MEG study performed on epileptic patients). In the second experiment (Exp 2), a native Finnish-speaking male (VJ) read a Finnish text (history of Soini, a small village in Finland, by Lasse Autio). Finnish was totally incomprehensible for the subjects who had never been exposed to this language; this speech stimulus had prosodic contours but no propositio-

nal content [Ischebeck et al., 2008] and was aimed to unravel the relationship of the observed reader-listener coupling to the understanding of speech content. Finally, a third experiment (Exp 3) was designed to find out whether the observed reader–listener coupling was specific to speech sounds: now the male French speaker (XDT) of Exp 1 "hummed" the text used in Exp 1. Exp 3 was conducted to provide vocal material that, in the same ecological situation as used in Exp 1 and Exp 2, preserved voice characteristics, such as $f0$ and the text rhythmic prosody, while removing speech content. Previous functional imaging studies have used recordings of natural and hummed speech to search for differences in the brain processing of the content or the more basic aspects of speech [Ischebeck et al., 2008]. During "humming," XDT uttered a vowel following as close as possible the speed and prosodic rhythmicity of the text (prosodic phrasing, punctuation, breathing, and natural pauses). The order of the experiments was randomized across subjects.

Data Acquisition

MEG signals were recorded at ULB-Hôpital Erasme with a whole-scalp-covering neuromagnetometer (Vectorview & Maxshield; Elekta Oy, Helsinki, Finland). Head position inside the MEG helmet was continuously monitored using four head-tracking coils. The locations of the coils with respect to anatomical fiducials were determined with an electromagnetic tracker (Fastrak, Polhemus, Colchester, VT). The passband for MEG and accelerometer signals was 0.1–330 Hz, and the sampling rate was 1 kHz. High-resolution 3D-T1 cerebral magnetic resonance images (MRI) were acquired on a 1.5 T MRI scan (Intera, Philips, The Netherlands).

Data Preprocessing

Continuous MEG data were first preprocessed off-line using the signal-space-separation method to suppress external interferences and correct for head movements [Taulu et al., 2005]. For frequency and coherence analysis, continuous MEG and accelerometer data were split into 2,048 ms epochs with 1,638-ms epoch overlap [Bortel and Sovka, 2007]. Such epoch length, leading to frequency resolution of 0.5 Hz (inverse of the epoch duration), is typical in coherence analyses [Baker, 2007; Bourguignon et al., 2011; Semmler and Nordstrom, 1999] as it offers a good compromise between frequency resolution and signal-to-noise ratio for computation of the coherence spectra. Indeed, an epoch length of 2,048 ms leads to frequency resolution of 0.5 Hz (inverse of the epoch duration), meaning that it allows to find coherence between accelerometer and MEG signals with a precision of 0.5 Hz. MEG epochs exceeding 3 pT (magnetometers) or 0.7 pT/cm (gradiometers) were excluded from further analysis to avoid contamination of the data by eye movements, muscle activity, or artifacts in

the MEG sensors. These steps led to more than 600 artifact-free epochs for each subject and condition.

We defined the $f0$ time-course as the Euclidian norm of the three accelerometer-signals band-passed around $f0$. Euclidian norm was taken in the spatial dimension, that is, the three acceleration time-courses were combined into one acceleration time-course. The $f0$ interval used to design the band-pass filter was determined by local minima surrounding the first peak of the smoothed acceleration power spectrum within the 70–270 Hz frequency band. The $f0$ time-course computed for each epoch was then used for coherence computation with MEG signals.

Coherence Analysis

The coherence is an extension of Pearson correlation coefficient to the frequency domain, quantifying the degree of coupling between two signals $x(t)$ and $y(t)$, providing a number between 0 (no linear dependency) and 1 (perfect linear dependency) for each frequency [Halliday et al., 1995]. Let $X_k(f)$ and $Y_k(f)$ be the Fourier transform of the kth segment of $x(t)$ and $y(t)$. By defining power spectra

$$P_{xx}(f) = \frac{1}{K}\sum_{k=1}^{K} X_k(f)X_k^*(f),$$

$$P_{yy}(f) = \frac{1}{K}\sum_{k=1}^{K} Y_k(f)Y_k^*(f)$$

and cross-spectrum

$$P_{xy}(f) = \frac{1}{K}\sum_{k=1}^{K} X_k(f)Y_k^*(f),$$

K being the number of segments used in formula (1–3), the coherence can be written as

$$\text{Coh}_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}.$$

Coherence was computed using formula (4) between the $f0$ time-course and artifact-free MEG epochs (i.e., sensor space) throughout the whole recorded frequency band (0.1–330 Hz due to hardware band-pass filter), yielding a coherence value for each possible combination of MEG sensors, frequencies, subjects, and conditions. According to formula (4), the coherence index at a given frequency does not depend on signal power at any other frequencies. Therefore, without *a priori* assumptions, all frequencies have the same potential to display significant coherence. Frequencies that had in all subjects significant coupling between the $f0$ time-course and sensor-level MEG signals were identified and defined as the frequencies of interest for coherent source analyses.

Analysis of Coherent Sources

Individual MRIs were first segmented using Freesurfer software (Martinos Center for Biomedical Imaging, MA). Then, the MEG forward model, comprising from 3711 to 4897 pairs of two orthogonal tangential current dipoles, placed on a homogeneous 7-mm-grid source space covering the whole brain, was computed using MNE suite (Martinos Center for Biomedical Imaging). To locate the coherent brain areas at the frequencies of interest (frequencies for which coherence was significant for all subjects), the $(N_b + 1 \times N_b + 1)$ cross-spectral density matrix $C(f)$ was computed between all possible combinations of MEG and $f0$ time-course signals, wherein N_b is the number of MEG signals (306 in Elekta Vectorview MEG device). $C(f)$ was computed only at frequencies of interest, using all artifact-free epochs of MEG and $f0$ time-course signals. On the basis of MEG forward model and cross-spectral density matrix, coherence maps were produced using the Dynamic Imaging of Coherent Sources (DICS), which uses spatial filter approach in frequency domain [Gross et al., 2001]. Coherence maps are parametric maps of coherence values, which were computed at individualized frequency bins corresponding to frequencies of interest and overlaid on subject's MRI. Separate coherence maps were computed for each possible combination of frequency of interest, subject, and condition. Both planar gradiometers and magnetometers were simultaneously used for inverse modeling after normalizing the MEG signals and forward model coefficients by 300 fT for magnetometer channels and 50 fT/cm for gradiometer channels.

Group Level Analysis

A 12-parameter affine transformation from individual MRIs to the standard Montreal Neurological Institute (MNI) brain was first computed using the spatial normalization algorithm implemented in Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK) and then applied to individual MRIs and coherence maps. This procedure generated normalized coherence maps in the MNI space for each subject, condition, and frequency of interest. To produce coherence maps at the group level, we computed the generalized f-mean across subjects of normalized maps, according to

$$f(\cdot) = \text{arctanh}(\sqrt{(\cdot)}),$$

namely, the Fisher z-transform of the square root. This procedure transforms the noise on the coherence into an approximately normally distributed noise [Rosenberg et al., 1989] resulting in an unbiased estimate of the group-level mean coherence. In addition, this averaging procedure avoids an excessive contribution from a single subject to the group analysis by using a function that is concave and monotonously increasing for typical coherence values.

The group level analysis yielded a coherence map for each combination of frequency of interest and condition.

Statistical Analyses

Coherence significance

Simulated data were used to assess the threshold for statistical significance of coherence values in single subject's sensor space and group-level coherence maps. This approach overcomes the multiple-comparison issue, which has no straightforward analytical solution when dealing with highly dependent time series.

As the Kolmogorov–Smirnov test was not able to reject the null hypothesis of gaussianity of the Fourier coefficients of the $f0$ time courses at the frequencies of interest, the coefficients' statistical distribution was assumed Gaussian. Autocorrelation in time was not modeled in the simulated $f0$ time-course, because we did not find significant dependency between adjacent and disjoint $f0$ time-course epochs at the frequencies of interest. To assess the statistical significance of our results, subject-level coherence values in the sensor space and group-level coherence maps were computed using real MEG signals and 10,000 randomly generated $f0$ time-course signals. Maximal coherence value(s) were then extracted for each simulation to compute the cumulative density function of the maximal coherence value(s) occurring due to stochastic matching between $f0$ time-course and MEG sensor or source signals. The coherence thresholds at $P < 0.05$, corrected for multiple comparisons, were then evaluated as the 0.95 percentile of the corresponding cumulative density function.

Differences between Exps 1–3

Variance of the $f0$ time-course power at frequencies of interest was analyzed by means of one-way repeated-measures ANOVA with experimental condition (Exps 1–3) as within-subject variable (SPSS version 13.0, LEAD Technologies).

Coherence levels (maximal coherence over all sensors at the frequencies of interest) were analyzed by means of one-way repeated-measures ANOVA with experimental condition (Exps 1–3) as within-subject variable (SPSS version 13.0, LEAD technologies).

Group-level SPM analyses

To identify statistically significant group-level differences in coherent brain areas between Exps 1–2 and Exp 3, individual coherence maps were introduced in a flexible factorial design using SPM8. As the first step, t-contrasts were computed between each speech experiment (Exp1f, Exp1m, and Exp 2) and the humming experiment (Exp 3). Then, to search for common coherent brain areas in the speech experiments, as opposed to humming experiments, SPM conjunction analysis approach was applied to dis-
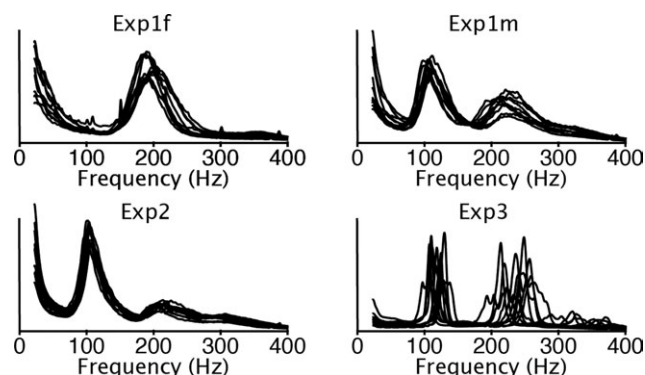
Figure 2.
Accelerometer power spectra. Power spectra of raw accelerometer signals representing the readers' voice that served as the stimuli for each participant (one line for each participant), in the four experiments. The traces illustrate the reproducibility of $f0$ across readers in each experiment. The power spectral analysis helped to determine the optimal band-pass filter for $f0$ interval computation.

close brain areas in which this effect was present in each and every contrast of interest [Friston et al., 1999].

RESULTS

Voice fundamental frequency

The accelerometer attached to the left side of the reader's throat accurately and reproducibly recorded the reader's voice $f0$ in all experiments (see Fig. 2). The mean $\pm$ SD $f0$ was 193 $\pm$ 7 Hz in Exp1f (female reader), 108 $\pm$ 5 Hz in Exp1m (male reader), 105 $\pm$ 3 Hz in Exp 2 (male reader), and 118 $\pm$ 10 Hz in Exp 3 (male reader). These values are consistent with the typical $f0$ values observed in adult males and females [Orlikoff, 1995].

Experiment 1: French text read by female or male speakers

In Exp1m and Exp1f, statistically significant right-hemisphere-dominant sensor-level coherence was found only at ~0.5 Hz in all subjects (coherence values from 0.052 to 0.307; $P_{corrected} < 0.05$; Fig. 3 and Table I). This 0.5 Hz frequency was therefore considered as the only frequency of interest to compute individual and group-level coherence maps for each experiment. Group-level coherence maps showed maximum coherence at the upper bank of the right posterior superior temporal sulcus (pSTS; coherence values 0.037 in Exp1f and 0.035 in Exp1m) and at the right posterior superior temporal gyrus (pSTG; coherence values 0.037 in Exp1f and 0.041 in Exp1m); the results were in both experiments statistically significant at $P_{corrected} < 0.05$ (see Fig. 4). In Exp1f, 8 of the 10 subjects had local maxima
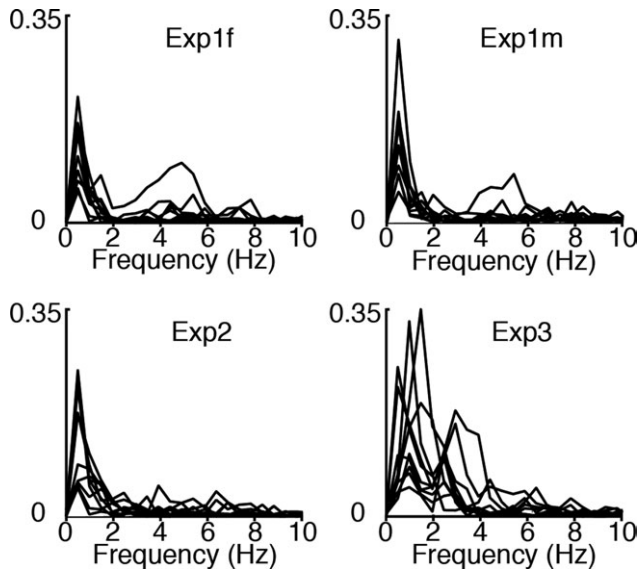
Figure 3.

Individual coherence spectra. Individual coherence spectra obtained in each condition for the MEG sensor showing the maximum coherence level. Significant coherence was observed in all subjects around 0.5 Hz in Exps 1 and 2. In Exp 3, significant coherence was found at 0.5 Hz in six subjects, at 1 Hz in two subjects, and at 1.5 Hz in two subjects.

at the right pSTS and 9 at the right pSTG. In Exp1m, 6 of the 10 subjects had local maxima at the right pSTS and 8 at the right pSTG. Additional statistically significant group-level maxima, although with weaker coherence, were found in Exp1f at the right inferior frontal gyrus (coherence value 0.017), right primary somatosensory cortex (S1; coherence value 0.014), and left parietal operculum (coherence value 0.019). Comparison between speech rhythm and phrasing, $f0$ time-courses, and MEG signals suggested a potential connection of the observed coherence to prosodic rhythm associated with natural pauses, punctuation marks, and breathing in the spoken text (see Fig. 5). Such prosodic pacing of reading indeed occurred on average at 0.58 Hz in Exp1f and at 0.45 Hz in Exp1m (averages across all subjects; the count of pauses of 100 ms, separated by at least one second, during the whole 300-s MEG recording, was 175 in Exp1f and 134 in Exp1m).

In four subjects (Exp1f and Exp1m), significant 4–6-Hz coherence was found at sensor level (coherence values 0.052–0.155; Fig. 3). Coherence maps at these frequencies revealed, in both experiments, maximum coherence at the supratemporal auditory cortex bilaterally, with no clear hemispheric dominance.

In none of the subjects did the coherence values at higher frequencies exceed the statistical threshold at the sensor level, and thus reconstructing the corresponding sources with DICS would have been improper.

Experiment 2: Finnish text read by a male speaker

In Exp 2, the sensor-level and source-level coherences were similar to those in Exp 1 (peaks at sensor level only at ∼0.5 Hz in all subjects; coherence values at sensor level 0.048–0.246; Fig. 3 and Table I), with strongest group-level coherence at the right pSTS (coherence value 0.039) and the right pSTG (coherence value 0.043; Fig. 4). Individual local coherence maxima occurred at the right pSTS in five subjects and at the right pSTG in nine subjects. In addition, significant group-level local maxima of less coherent sources were found in the left parietal operculum (coherence value 0.019) and left anterior STG (coherence value: 0.013). Comparison between speech rhythm and phrasing, $f0$ time-courses, and MEG signals again suggested a potential connection of the observed coherence to prosodic rhythm: pauses in the spoken text occurred on average at 0.42 Hz (on average 127 pauses during the MEG recording).

In two subjects, significant 4–6-Hz right-hemisphere-dominant coherence was found at the sensor level (coherence values 0.053–0.066; Fig. 3), with the maximum at the supratemporal auditory cortex bilaterally, with a clear right-hemisphere dominance.

As in Exp 1, coherence values at higher frequencies did not exceed the statistical threshold at the sensor level

Experiment 3: Text hummed by a male speaker

Significant right-hemisphere-predominant coherence peaking at 0.5 Hz in six subjects, at 1 Hz in two subjects, and at 1.5 Hz in two subjects was found at the sensor level (coherence values from 0.049 to 0.328, $P_{corrected} <$ 0.05; Fig. 3). However, the maximum coherence occurred, in all individuals and at the group level, at the right supratemporal auditory cortex (ACx; group-level

TABLE I. Coherence values in the sensor space for each subject and condition

	Exp1f	Exp1m	Exp 2	Exp3
Subject 1	0.0527	0.0818	0.0483	0.3278
Subject 2	0.0877	0.0804	0.1714	0.1492
Subject 3	0.1494	0.1302	0.1746	0.1930
Subject 4	0.0703	0.0518	0.0874	0.0833
Subject 5	0.1622	0.1647	0.2175	0.0970
Subject 6	0.1680	0.1862	0.0576	0.1461
Subject 7	0.1120	0.1255	0.0601	0.0491
Subject 8	0.0579	0.0960	0.0510	0.1055
Subject 9	0.2116	0.3072	0.2456	0.1563
Subject 10	0.1558	0.1802	0.2385	0.0724
Mean	0.1228	0.1404	0.1352	0.1380
Min	0.0527	0.0518	0.0483	0.0491
Max	0.2116	0.3072	0.2456	0.3278

ANOVA repeated measures: $P = 0.606$.
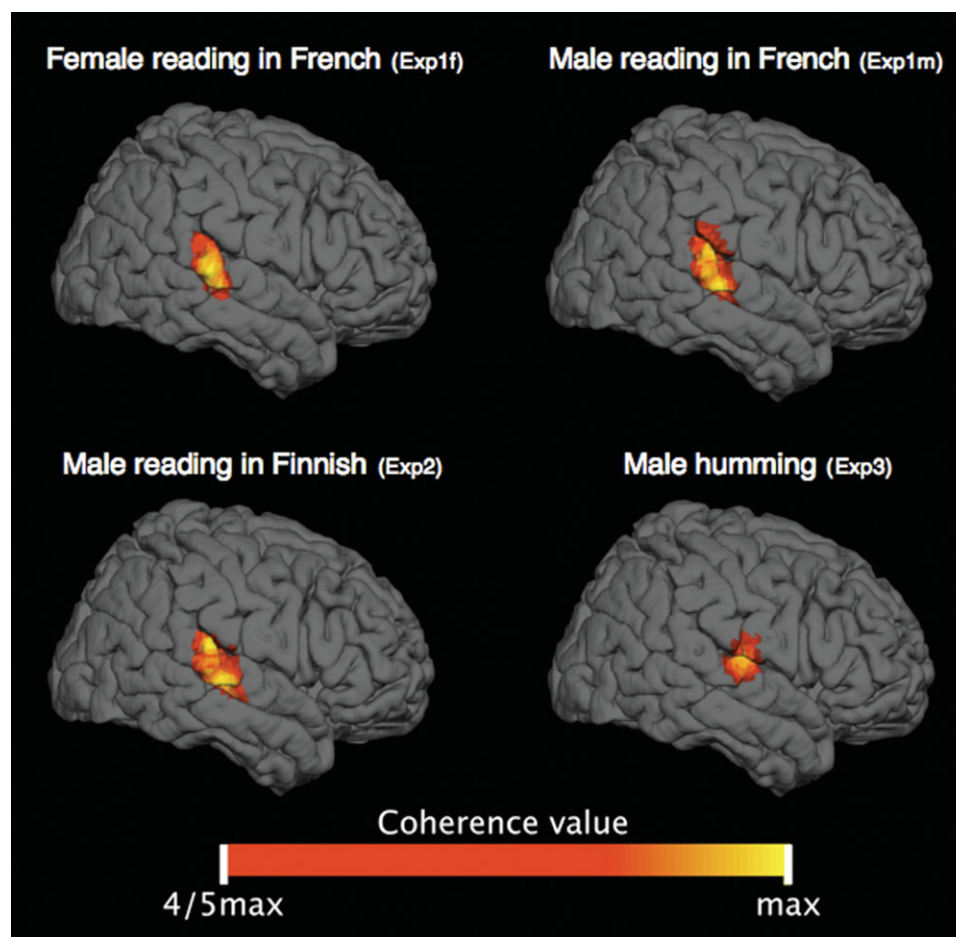
Figure 4.
Group-level coherence maps. Group-level coherence maps computed at 0.5 Hz in each condition and displayed for the right hemisphere. In Exps 1 and 2 (top and bottom left), the maximum coherence occurred at the superior bank of the right posterior superior temporal sulcus and the posterior superior temporal gyrus. In Exp 3 (bottom right), the maximum coherence was located at the right supratemporal auditory cortex.

coherence value: 0.034; Fig. 4). Weaker but still statistically significant local group-level maxima were found at the left ACx (coherence value 0.013) and parietal operculum (coherence value 0.014). Comparison between speech rhythm and phrasing, $f0$ time-courses, and MEG signals suggested a potential connection of the observed coherence to prosodic rhythm associated to pauses that during humming occurred on average at 0.44 Hz (on average 133 pauses during the recording). Compared with the other experiments, this number of pauses did not differ significantly across subjects (paired t-test; Exp1f vs. Exp3, $P = 0.09$; Exp1m vs. Exp3, $P = 0.095$; Exp2 vs. Exp3, $P = 0.77$).

In four subjects, significant right-hemisphere-dominant 4–6-Hz coherence was found at sensor level (coherence values 0.057–0.148; Fig. 3); the maximum coherence occurred at the supratemporal auditory cortex bilaterally, with a clear right-hemisphere dominance.
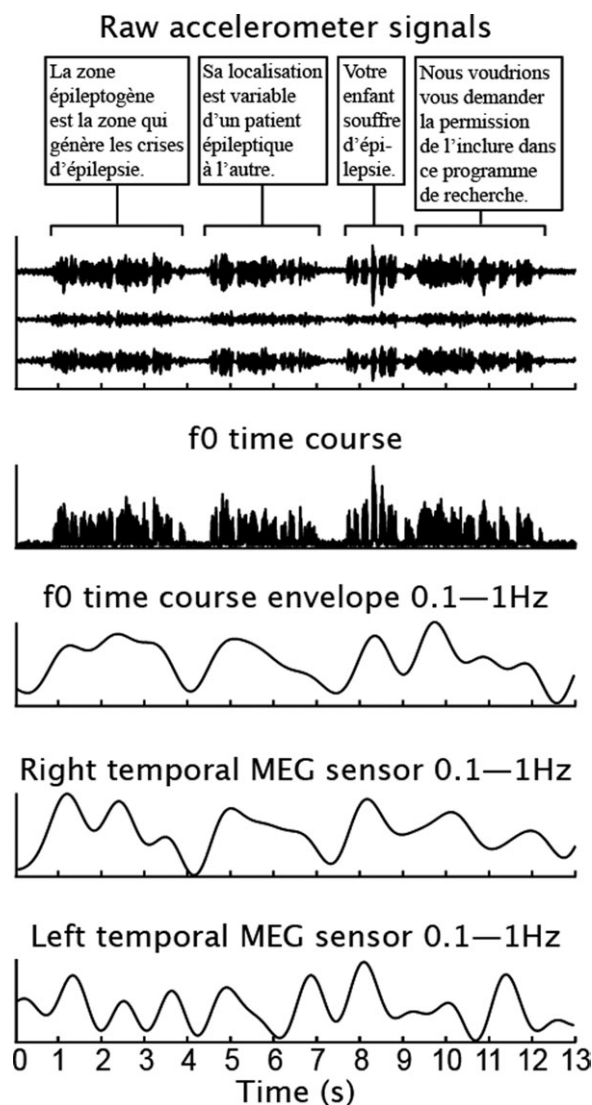
As in previous experiments, coherence values at higher frequencies did not exceed the statistical threshold in sensor-level analysis.

Differences between Exps 1–3

The variance of the $f0$ time-course power at 0.5 Hz did not differ between Exps 1–3 [$F(3,36) = 0.72$, $P = 0.55$].

The levels of sensor-level 0.5-Hz coherence did not differ between Exps 1–3 ($P = 0.61$).

Raw accelerometer signals

| La zone épileptogène est la zone qui génère les crises d'épilepsie. | Sa localisation est variable d'un patient épileptique à l'autre. | Votre enfant souffre d'épilepsie. | Nous voudrions vous demander la permission de l'inclure dans ce programme de recherche. |

f0 time course

f0 time course envelope 0.1—1Hz

Right temporal MEG sensor 0.1—1Hz

Left temporal MEG sensor 0.1—1Hz

0 1 2 3 4 5 6 7 8 9 10 11 12 13
Time (s)

Figure 5.

Comparison between speech rhythm, phrasing, $f0$ time-course, and magnetoencephalographic (MEG) signals for Subject 9 in Exp1m, suggesting a potential connection of the observed coherence to prosodic phrasing rhythmicity associated with pauses in the spoken text, such as those linked to punctuation marks (commas and periods). The slow fluctuations of the $f0$ envelope and MEG signals are more correlated for right temporal MEG sensors than for left temporal MEG sensors, explaining why coherence between $f0$ envelope and MEG signals was maximal at right temporal cortical areas.

Group-level SPM analyses

For the observed coherence at 0.5 Hz, the SPM conjunction analysis of speech versus nonspeech experiments indicated that signals at the right pSTS [68 −38 18] and pSTG

[58 −38 24] fluctuated more coherently with the reader's $f0$ time-course during both comprehensible (Exp 1) and uncomprehensible (Exp 2) speech than during humming (Exp 3) (see Fig. 6).

DISCUSSION

Using a novel corticovocal coherence method, we found significant coupling in all subjects at about 0.5 Hz between the reader's voice $f0$ time-course and the listener's cortical MEG during natural listening. The observed coupling demonstrates the existence of common fluctuations in the reader's vocal and the listener's brain signals occurring on average every 2 s. The brain regions showing the maximum coherence were different for speech and nonspeech vocal sounds, but all were lateralized toward the right temporal lobe. In some subjects, we also found significant 4–6 Hz coherence in the supratemporal auditory cortex bilaterally.

The coherence phenomenon

The very similar coherence levels between reader's voice $f0$ time course and the listener's MEG signals for both comprehensible (Exp 1) and incomprehensible (Exp 2) speech, as well as for humming (Exp 3), raise the key issue of the underlying neurophysiological mechanisms. Activity of right-hemisphere auditory areas has been suggested to follow the voice envelope, as defined by sound onsets and offsets [Abrams et al., 2008; Ahissar et al., 2001]. Accordingly, previous animal and human neuroimaging studies have suggested that right-hemisphere auditory cortex preferentially encodes slow (4–8 Hz) temporal features of syllable patterns ("envelope peak-tracking units") [Abrams et al., 2008]. Our finding of 4–6-Hz coherence in some subjects supports this hypothesis, but our most prominent results extend the findings to phrasal or sentence patterns, with envelope frequencies around 0.5 Hz. Interestingly, while the level of coherence was similar at the sensor level for speech and nonspeech sounds, the brain regions showing the maximum coherence differed; still all patterns were lateralized towards the right hemisphere. This finding suggests that tracking the 0.5-Hz acoustical envelopes involves in the right temporal cortex different neuronal circuitries for speech and non-speech vocal sounds.

The coherence values observed at sensor and source levels ranged from 0.03 to 0.3. These values may seem surprisingly low for a relevant neurophysiological phenomenon, even though statistically significant after correction for multiple comparisons. Physiologically, relevant coherence values below 0.1 have been repeatedly reported in studies investigating the coherence between brain activity and electromyographic activity or movement frequency [Conway et al., 1995; Gross et al., 2005; Pohja et al., 2005; Pollok et al., 2004; Salenius et al., 1997]. As the likely triggering event of the corticovocal coherence—the periodic variation of the speech envelope due to pauses and prosodic
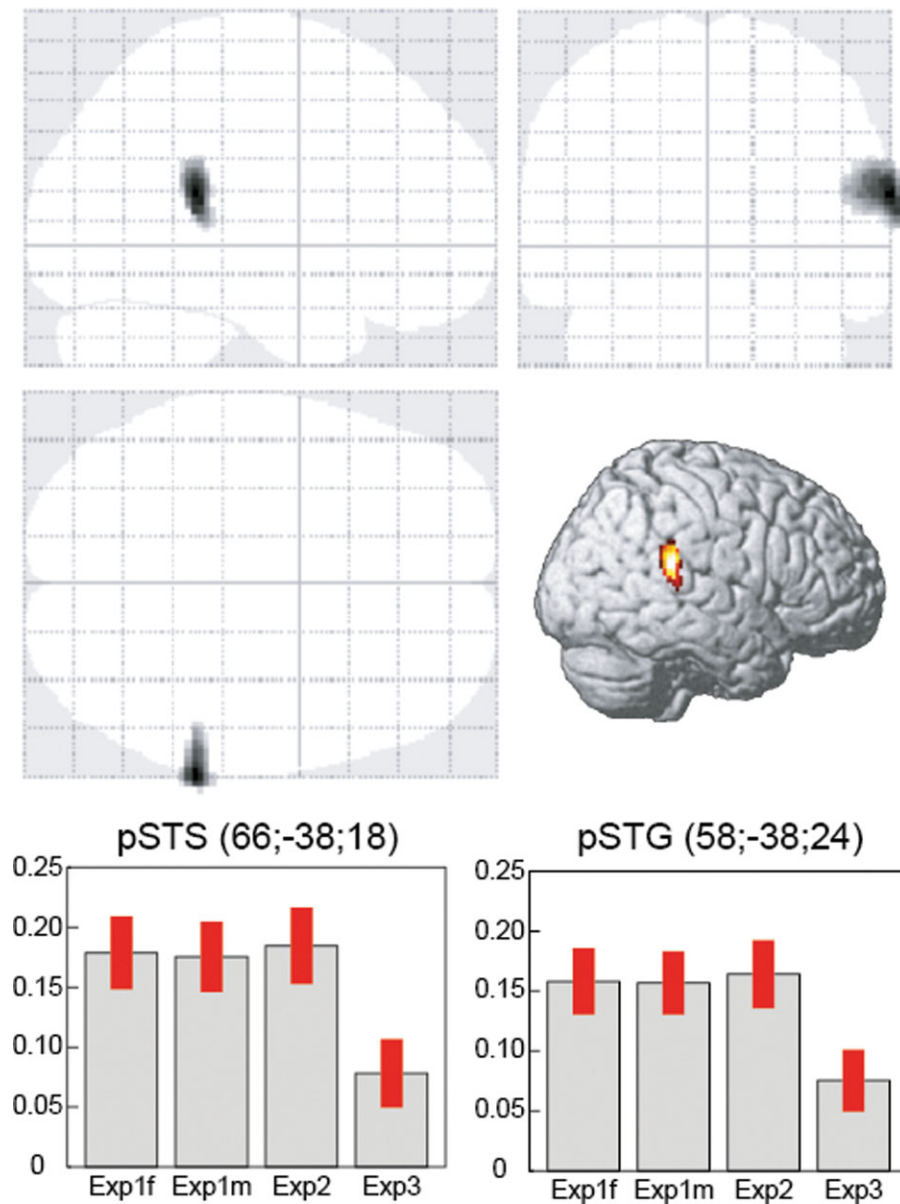
Figure 6.

Statistical parametric mapping conjunction analysis. Results of SPM conjunction analysis of comprehensible (French) and incomprehensible (Finnish) speech versus humming. *Top and middle left*: Results on the default SPM glass-brain. *Middle right*: Results overlayed on the SPM-template 3D-brain rendering. Statistical maps are displayed at an uncorrected threshold of 0.001. *Bottom*: Mean and 90% confidence interval of the *f*-transformed coherence values (according to *f* as defined in the Methods section) for each experiment at the right pSTS and pSTG.

pacing—varies in its rhythm during natural speech (despite the observed rather similar mean number of pauses in different conditions), the coherence at the mean frequency of the rhythm is naturally diminished although the most coherent brain regions stay stable. Considering the brain areas involved in the corticovocal coherence described in this work, the phenomenon appears physiologically meaningful despite the small coherence values.

Speech-sensitive brain areas for reader-listener coupling

Our findings—strong coherence at the right pSTS and pSTG for speech but only at the right auditory cortex for humming—suggest strong speech-sensitivity, even though not strict speech-specificity, for the corticovocal coherence. Although coherence was peaking at 1 Hz (two subjects) and at 1.5 Hz (two other subjects) in Exp 3, it did not affect the brain region showing significant coupling with the $f0$ time-course.

The observed differences in the locations of coherent brain areas between speech and nonspeech experiments support previous functional imaging findings on different neuronal networks processing naturally spoken versus hummed sentences; for example, the middle and posterior parts of the STG are activated more strongly for natural than hummed speech [Ischebeck et al., 2008]. The right pSTS and pSTG are well-known key nodes of the speech-processing network [Hickok and Poeppel, 2007]. They are involved in the integration of spectral and temporal modulations of human voice [Belin et al., 2004; Kriegstein and Giraud, 2004; Lattner et al., 2005; Warren et al., 2005] and respond more strongly to verbal than nonverbal vocalizations [Belin et al., 2000]. Both pSTS and pSTG have been further implicated, together with their left-hemisphere homologous regions, in sentence processing, such as the integration of lexical-semantic and syntactic information during sentence comprehension [Friederici et al., 2009]. Finally, the right pSTG plays a particular role in the processing of slow prosodic modulations in speech sounds [Meyer et al., 2002].

Increasing evidence supports the involvement of the right pSTS also in interindividual communicative behavior, for example, contributing to inferences about communicative intentions and to generation of communicative actions [Noordzij et al., 2009]. In this context, the strong reader–listener coupling at the right pSTS, observed in our listening settings, might also represent a neuronal correlate of the prediction or recognition, in the listener, of a forthcoming communication intention on the reader's side.

Temporal windows for auditory processing

Successful speech comprehension involves segmentation of speech signals on various timescales or temporal windows (TWs), so that the spectrotemporal information is concurrently extracted within separate neural streams for speech perception and then integrated in brain regions involved in subsequent lexical computations for speech comprehension [Abrams et al., 2008; Friederici, 2002; Hickok and Poeppel, 2007; Rosen, 1992]. Although the spectrotemporal receptive fields in primary auditory cortices appear to be symmetric between the hemispheres, evidence based on functional neuroimaging favors asymmetric processing in nonprimary auditory areas. For example, speech and nonspeech signals are analyzed in different TWs ranging from around 20 to >1,000 ms [Poeppel et al., 2008]. Two timescales have been investigated in depth: a 20–50 ms TW in left nonprimary auditory areas and a 150–300 ms TW in right nonprimary auditory areas [Abrams et al., 2008; Boemio et al., 2005; Luo and Poeppel, 2007; Poeppel et al., 2008]. In the context of speech processing, the 20–50 ms TW corresponds to durations of phonemes, while the 150–300 ms TW corresponds to the duration of syllables and timescales of prosodic phenomena [Poeppel et al., 2008]. Longer TWs, at the level of the second(s), are supposed to be involved in sentence-level information [Poeppel et al., 2008]. The coherence at 0.5 Hz between the reader's vocal and the listener's brain signals in our study would correspond to a TW of about 2 s. Analysis of the sound envelopes suggests a potential connection of this timescale to the pace of prosodic rhythm associated with pauses or boundaries within and between sentences of the spoken or hummed text. This hypothesis is in line with findings that in speech perception, the $f0$ envelope mainly transfers information on the sentence structure, such as phrase boundaries [Lattner et al., 2005]. Thus, it is understandable why we observed significant coherence in all subjects for TWs corresponding to sentence or phrasal levels and not to those corresponding to syllable or phoneme levels, both of which are more carried by formants [Lattner et al., 2005]. These data therefore suggest that the right hemisphere is specifically involved in the processing of vocal sounds on timescales extending from hundreds of milliseconds to several seconds. These results further support the hypothesis of hemispheric asymmetry in processing of speech and nonspeech sounds in nonprimary auditory areas, now in time windows up to 2 s.

Importantly, we found that reader–listener coupling in the pSTS and pSTG during all speech experiments, independently whether the subject understood the language. Thus, the observed coupling in the pSTS and pSTG appears to be related to prelinguistic, acoustic processing of speech inputs [Abrams et al., 2008]. Nevertheless, behavioral and electrophysiological evidence suggests that prosodic cues, such as pauses or boundaries in spoken language directly influence syntactic processing, and therefore convey critical information helping to understand the meaning of sentences [Friederici, 2004; Steinhauer and Friederici, 2001; Steinhauer et al., 1999]. Correspondingly, some authors consider that the prosodic representation of sentences is involved in the understanding of spoken language by supplying the basic framework that allows us to hold an auditory linguistic sequence in working memory during further processing [Frazier et al., 2006; Friederici, 2004].

An additional aspect to consider in the context of our study is the importance of pauses and prosodic boundaries in interactional speech communication to guide the conversation flow between the participants [Wilson and Wilson, 2005]. Indeed, conversational partners typically communicate turn by turn, each participant taking alternatively the speaker and listener's role. Pauses and silences

in continuous speech represent crucial cues for turn-taking transitions, and the between-speakers pauses during sustained speech interaction typically last a few hundreds milliseconds [Wilson and Wilson, 2005]. Although our subjects were only passively listening to speech, the observed reader–listener coupling might actually relate to neuronal processes involved in the recognition of relevant transitions in the reader's voice, facilitating the listener to eventually initiate speech during interactional verbal communication. The right pSTS might play a crucial role in such interindividual communicative behavior. Further investigations using interactional verbal paradigms are clearly needed to confirm this hypothesis.

Measuring corticovocal coherence

The simple and ecologically valid corticovocal coherence method presented in this study may open new possibilities for future time-sensitive investigations of processing of natural continuous speech. In particular, it could be applied to investigate the coherent neuronal networks between two individual brains engaged in natural verbal communication. The method might also be applied to study development of language function, both in healthy children and under various pathological conditions.

CONCLUSIONS

We demonstrated the existence of right-hemisphere dominant significant coupling at about 0.5 Hz between the fundamental frequency of the reader's voice and the listener's cortical activity, as followed by MEG, during listening to natural continuous speech. The observed coupling indicates common fluctuations in the reader's voice and the listener's brain activity occurring on average every 2 s, likely related to variations of acoustical envelopes in association of prosodic rhythmicity. The brain regions showing the maximum coherence for speech sounds, the right pSTS and pSTG, are among the key nodes of brain circuitries supporting speech processing and interindividual communicative behavior. Our findings also bring new support to the hypothesis of hemispheric asymmetry in processing speech and nonspeech signals extending over different timescales. At a more general level, the corticovocal coherence method developed in this study opens new perspectives for investigation of speech processing in ecologically valid listening situations.

ACKNOWLEDGMENTS

REFERENCES

Abrams DA, Nicol T, Zecker S, Kraus N (2008): Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J Neurosci 28:3958–3965.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001): Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci USA 98:13367–13372.

Aiken SJ, Picton TW (2008): Human cortical responses to the speech envelope. Ear Hear 29:139–157.

Baker SN (2007): Oscillatory interactions between sensorimotor cortex and the periphery. Curr Opin Neurobiol 17:649–655.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belin P, Fecteau S, Bedard C (2004): Thinking the voice: Neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Boemio A, Fromm S, Braun A, Poeppel D (2005): Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395.

Bortel R, Sovka P (2007): Approximation of statistical distribution of magnitude squared coherence estimated with segment overlapping. Signal Process 87:1100–1117.

Bourguignon M, De Tiège X, de Beeck MO, Pirotte B, Van Bogaert P, Goldman S, Hari R, Jousmäki V (2011): Functional motor-cortex mapping using corticokinematic coherence. Neuroimage 55:1475–1479.

Conway BA, Halliday DM, Farmer SF, Shahani U, Maas P, Weir AI, Rosenberg JR (1995): Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man. J Physiol 489:917–924.

Frazier L, Carlson K, Clifton C Jr. (2006): Prosodic phrasing is central to language comprehension. Trends Cogn Sci 10:244–249.

Friederici AD (2002): Towards a neural basis of auditory sentence processing. Trends Cogn Sci 6:78–84.

Friederici AD (2004): Processing local transitions versus long-distance syntactic hierarchies. Trends Cogn Sci 8:245–247.

Friederici AD, Makuuchi M, Bahlmann J (2009): The role of the posterior superior temporal cortex in sentence comprehension. Neuroreport 20:563–568.

Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ (1999): Multisubject fMRI studies and conjunction analyses. Neuroimage 10:385–396.

Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R (2001): Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci USA 98:694–699.

Gross J, Pollok B, Dirks M, Timmermann L, Butz M, Schnitzler A (2005): Task-dependent oscillations during unimanual and bimanual movements in the human primary motor cortex and SMA studied with magnetoencephalography. Neuroimage 26:91–98.

Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF (1995): A framework for the analysis of mixed time series/point process data–theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog Biophys Mol Biol 64:237–278.

Hari R, Kujala MV (2009): Brain basis of human social interaction: From concepts to brain imaging. Physiol Rev 89:453–479.

Hickok G, Poeppel D (2007): The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hillman RE, Heaton JT, Masaki A, Zeitels SM, Cheyne HA (2006): Ambulatory monitoring of disordered voices. Ann Otol Rhinol Laryngol 115:795–801.

Ischebeck AK, Friederici AD, Alter K (2008): Processing prosodic boundaries in natural and hummed speech: An FMRI study. Cereb Cortex 18:541–552.

Jerbi K, Lachaux JP, N'Diaye K, Pantazis D, Leahy RM, Garnero L, Baillet S (2007): Coherent neural representation of hand speed in humans revealed by MEG imaging. Proc Natl Acad Sci USA 104:7676–7681.

Kriegstein KV, Giraud AL (2004): Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955.

Lattner S, Meyer ME, Friederici AD (2005): Voice perception: Sex, pitch, and the right hemisphere. Hum Brain Mapp 24:11–20.

Lindström F, Ren K, Li H, Waye KP (2009): Comparison of two methods of voice activity detection in field studies. J Speech Lang Hear Res 52:1658–1663.

Lindström F, Ohlsson AC, Sjoholm J, Waye KP (2010): Mean F0 values obtained through standard phrase pronunciation compared with values obtained from the normal work environment: A study on teacher and child voices performed in a preschool environment. J Voice 24:319–323.

Luo H, Poeppel D (2007): Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010.

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002): FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum Brain Mapp 17:73–88.

Noordzij ML, Newman-Norlund SE, de Ruiter JP, Hagoort P, Levinson SC, Toni I (2009): Brain mechanisms underlying human communication. Front Hum Neurosci 3:14.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA III, Brugge JF (2009): Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564–15574.

Orlikoff RF (1995): Vocal stability and vocal tract configuration: An acoustic and electroglottographic investigation. J Voice 9:173–181.

Poeppel D, Idsardi WJ, van Wassenhove V (2008): Speech perception at the interface of neurobiology and linguistics. Philos Trans R Soc Lond B Biol Sci 363:1071–1086.

Pohja M, Salenius S, Hari R (2005): Reproducibility of cortex-muscle coherence. Neuroimage 26:764–770.

Pollok B, Gross J, Dirks M, Timmermann L, Schnitzler A (2004): The cerebral oscillatory network of voluntary tremor. J Physiol 554:871–88.

Rosen S (1992): Temporal information in speech: Acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373.

Rosenberg JR, Amjad AM, Breeze P, Brillinger DR, Halliday DM (1989): The Fourier approach to the identification of functional coupling between neuronal spike trains. Prog Biophys Mol Biol 53:1–31.

Sabbagh MA (1999): Communicative intentions and language: Evidence from right-hemisphere damage and autism. Brain Lang 70:29–69.

Salenius S, Portin K, Kajola M, Salmelin R, Hari R (1997): Cortical control of human motoneuron firing during isometric contraction. J Neurophysiol 77:3401–3405.

Scott SK (2008): Voice processing in monkey and human brains. Trends Cogn Sci 12:323–325.

Semmler JG, Nordstrom MA (1999): A comparison of cross-correlation and surface EMG techniques used to quantify motor unit synchronization in humans. J Neurosci Methods 90:47–55.

Steinhauer K, Friederici AD (2001): Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. J Psycholinguist Res 30:267–295.

Steinhauer K, Alter K, Friederici AD (1999): Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci 2:191–196.

Stephens GJ, Silbert LJ, Hasson U (2010): Speaker-listener neural coupling underlies successful communication. Proc Natl Acad Sci USA 107:14425–14430.

Suppes P, Lu ZL, Han B (1997): Brain wave recognition of words. Proc Natl Acad Sci USA 94:14965–14969.

Suppes P, Han B, Lu ZL (1998): Brain-wave recognition of sentences. Proc Natl Acad Sci USA 95:15861–15966.

Taulu S, Simola J, Kajola M (2005): Applications of the signal space separation method. IEEE Trans Sign Proc 53:3359–3372.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003): Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17:48–55.

von Kriegstein K, Smith DR, Patterson RD, Kiebel SJ, Griffiths TD (2010): How the human brain recognizes speech in the context of changing speakers. J Neurosci 30:629–638.

Warren JD, Jennings AR, Griffiths TD (2005): Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24:1052–1057.

Wilson M, Wilson TP (2005): An oscillator model of the timing of turn-taking. Psychon Bull Rev 12:957–968.